Sensor network data analysis at different levels of granularity

Susanne Bleisch
Department of
Infrastructure Engineering,
University of Melbourne
Melbourne, Australia
susanne.bleisch@unimelb.edu.au

Matt Duckham
Department of
Infrastructure Engineering,
University of Melbourne
Melbourne, Australia
matt@duckham.org

Abstract

The analysis of sensor network data may reveal insights at different levels of granularity. Based on the example of analysing fish movement data in Murray River in Australia, it is shown how the use of different levels of granularity allow the aggregation of otherwise sparse data and find insights into fish movement behaviour. The approach is then conceptually extended to more complex network structures where animals are not restricted to linear movement by using a graph-theoretic perspective for landscape connectivity.

Keywords: sensor network data, analysis, granularity, animal movement.

1 Introduction

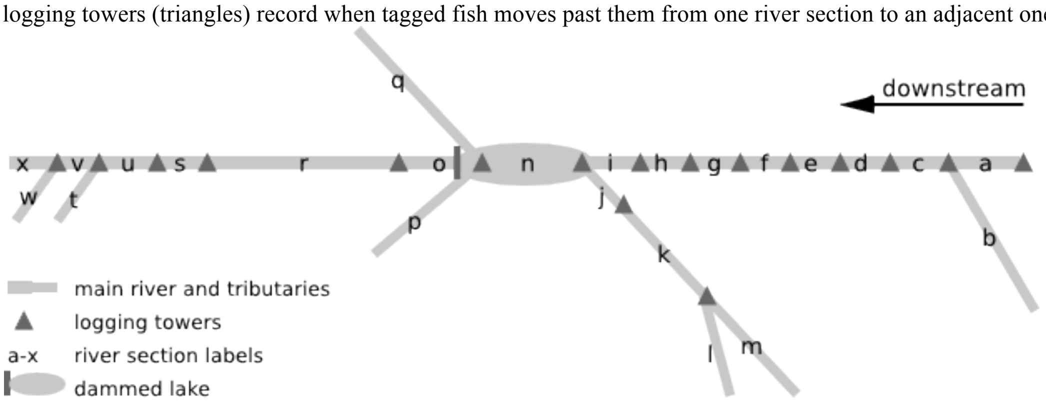
A challenge in analysing sensor network data is that the data may be collected with varying granularity. Additionally, the collected data may also reveal information or promote insight at different levels of data selection and aggregation (semantic granularity, Hornsby cited in [1]). Various methods, from statistics to visualisation, can be used for the data analysis but the results are often dependent on the chosen level of granularity. This poster proposes an approach to use different levels of granularity to help overcoming scarcity of data even in seemingly large data sets based on the example of analysing fish movement data in a river. The approach is then conceptually extended to more generic network structures.

2 Analysing linearly restricted data

A current project in the Murray River in South Eastern Australia is monitoring the native fish population following fish habitat improvements in parts of the river. 18 logging towers segment the river into different sections (labelled 'a' to 'x') and record the movements of in total (during several years) more than 1000 tagged fish between those river sections (Figure 1). Fish ecologists have analysed the data set to evaluate the success of different management intervention. However, the rich data set may be able to reveal more information about common and unusual fish behaviour.

The fish data set contains daily recordings of fish moving between river sections. Thus, the basic movement event is the change from one river section to another, most often adjacent, river section at a specific day. This could be fish moving from section f to section g (f>g). However, when interested in common fish behaviour, this characterization of fish movement, containing references to specific sections, may be too precise. A fish moving e>f or d>e, for example, might both be generalised to a single type of movement: moving one section downstream. Similarly, the fish ecologists are interested in sections where fish habitats may be different, such as in the dammed lake or sections of the tributaries (cf. Figure 1) and whether fish move to or out of such sections. Looking at the individual sections may reveal more detail, but for a more general picture it is useful to generalise the specific movements on a more generic level, such as all movements from the main river and the tributaries to the lake. Thus, each section change (each basic fish movement) gets assigned labels at different levels of granularity, including: the exact

Figure 1: Schematic drawing of the monitored part of the Murray River including the dammed lake Mulwala (section n). The logging towers (triangles) record when tagged fish moves past them from one river section to an adjacent one.



section labels; the distance (one or more sections); the direction (up- and downstream); and moving to/from sections of specific interest. The movement i>n would thus be labelled 'i to n' but also 'one river section downstream' and 'moving from the main river to the dammed lake' (cf. Figure 1).

In one project [2] sequence mining was used to find interesting fish movement patterns, such as fish moving a long distance upstream for spawning. In the original description of the basic events, moving a long distance upstream could mean, for example, r>h, n>d or r>c. A potentially long list of different events would need to be extracted and aggregated to get an overview of long upstream movements. The more generic labels for the fish movements like 'upstream' simplify the sequence mining for spawning movements upstream.

3 Analysing networks of movement data

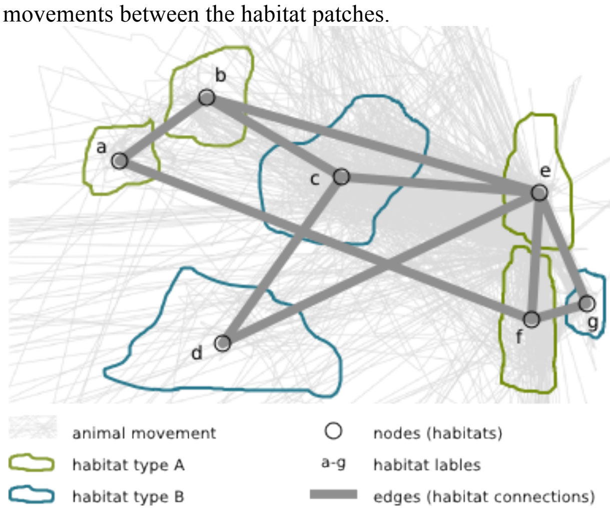
The discussed approach serves well for a mainly linear geographic structure, such as a river, with clear directions (e.g. up- and downstream). However, adopting a graph-theoretic perspective for landscape connectivity [3] allows this approach to be generalized to areal structures or networks (potentially also in three dimensions) where movement is not restricted linearly. In such settings, nodes could, for example, be individual habitat patches and edges represent the summarised movement of animals between those habitat patches [4]. This approach is illustrated in Figure 2 through a fictitious landscape consisting of several habitat patches of hypothetical interest and animal movements between them. By describing these movements at different levels of granularity, such as in the fish movements in the example above, the detection of more common movement patterns between the different habitat patches can be facilitated (for example in Figure 2, moving b>c or e>d could more generally be described as moving 'habitat type A'>'habitat type B'). Additionally, using different levels of granularity may also facilitate visualisation and visual analysis of the data or data mining results. Especially for large and/or dense data set, aggregation at different levels [e.g. 5] can help visualisations be more easily understood or emphasize specific patterns in the data set.

4 Discussion and Outlook

The abstraction of a river or a landscape to interesting sections or habitat patches needs the input of domain experts. Without such input the results of the analysis are likely to be meaningless. Similarly, the definition of the aggregation levels needs expert knowledge. However, while manual input is needed, the definitions can be encoded in rules and automatically applied to data sets.

In case where different levels of aggregation can be defined this seems to be an approach able to help solve the common sparse data problem. In the fish movement data case study, we were able to show that fish movements between specific river sections are too sparse to yield meaningful insights about fish behaviour. However, aggregating the data to movement in general and especially up- and downstream movement we were able to gain meaningful insights into fish movement in the Murray River. Whether these finding are valid also for more general networks, as discussed in section 3, will be tested in further studies.

Figure 2: Clipping of animal movement trajectories overlaid with fictitious habitat patches of type A (green) and type B (blue), which are of hypothetical interest. The patches are reduced to nodes (circles) that are connected with edges (thick grey lines). The edges represent the summarized animal movements between the habitat patches.



Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve this abstract.

References

[1] F. Fonseca, M. Egenhofer, C. Davis, and G. Camara, "Semantic Granularity in Ontology-Driven Geographic Information Systems," *AMAI Annals of Mathematics and Artificial Intelligence - Special Issue on Spatial and Temporal Granularity*, 36(1–2); 121–151, 2002.

[2] S. Bleisch, M. Duckham, P. Laube, and J. Lyon, "Identifying candidate causal relationships in fish movement patterns," in *Proc. 21st GIS Research UK (GISRUK)*, 2013.

[3] D. Urban and T. Keitt, "Landscape Connectivity: A Graph-Theoretic Perspective," *Ecology*, 82(5):1205–1218, 2001.

[4] Ö. Bodin and J. Norberg, "A network approach for analyzing spatially structured populations in fragmented landscape," *Landscape Ecology*, 22:31–44, 2007.

[5] G. Andrienko and N. Andrienko, "A General Framework for Using Aggregation in Visual Exploration of Movement Data," *The Cartographic Journal*, 47(1):22–40, 2010.